

Using Mission Essential MOEs/MOPs for Evaluating Effectiveness of Distributed Mission Training

Frank C. Gentner, T. Cliff Tiller, and Paul H. Cunningham
University of Dayton Research Institute (UDRI) Human Factors Group
300 College Park, Dayton, OH 45469-0157, U.S.A.

Winston Bennett, Jr.
Air Force Research Laboratory (AFRL) Human Effectiveness Directorate
Warfighter Training Research Division
6030 S. Kent, Mesa, AZ 85212-6004, U.S.A.

ABSTRACT

Demonstrating the mission relevance of advanced training and rehearsal systems and their focus on training and evaluating warfighter needs is best achieved with objective metrics that can highlight mission performance changes. However, for a variety of reasons, it has historically been much easier to evaluate training in the traditional ways, that is, focusing on student evaluations and end-of-course tests as opposed to examining on-the-job behaviors and organizational or mission success. Student evaluations and learning tests are easy to implement but may or may not be explicitly tied to the overall training objectives. Further, these assessments do not provide any indication of the impact of training on job performance or mission effectiveness. Critical reviews found lack of an integrated system for measuring and assessing training performance, over-reliance on subjective measures of performance, and a shortage of valid, reliable, quantitative performance measures of training effectiveness. This paper highlights initial research and data collected to develop an Aircrew Measure of Effectiveness (MOE)/Performance (MOP) Hierarchical Taxonomy capable of assisting training and mission evaluators. The paper details our approach and provides data on sample mission task MOE/MOP decompositions to illustrate how a taxonomic approach can help diagnose actual aircrew mission performance of both individuals and teams. While this approach shows much promise, many technical obstacles need to be overcome before it can be completed and used routinely in an automated form. We highlight and discuss these technical challenges, propose solutions, and provide an agenda for needed research. Implications and potential future applications of the approach are discussed.

AUTHORS

Mr. Frank C. Gentner, as senior technical analyst, manages Human System Integration (HSI)-related projects at the UDRI Human Factors Group. Recent projects include analyses of competencies of simulation training observer/controllers, automated mission planning system human factors and training issues, developing operational measures of effectiveness for warfighter performance, and a major update to MIL-HDBK-46855A, *Human Engineering Program Tasks and Procedures*. During his 23-year military career, Lt Col Gentner (Ret.) served in the Air Force (AF) model HSI organization, supporting Aeronautical Systems Division system program offices. He led AF and OSD activities in training requirements, development, and evaluation functions. He holds a Bachelor of Psychology and Masters in Rehabilitation from the University of Florida.

Mr. T. Cliff Tiller is a Human Factors Engineer (HFE) in the UDRI Human Factors Group. As 15-year AF F-16 pilot, he brings a wealth of subject matter expertise to the UDRI team. He has contributed to the development of the Aircrew MOE Taxonomy and the update to MIL-STD-1472. Cliff holds an MS in HFE as well as an MS in Aeronautics.

Mr. Paul H. Cunningham, as a senior technical analyst with UDRI's Human Factors Group, is responsible for managing projects like the AF validation of the Manpower Personnel and Training Decision Support System (MPT DSS) for the Joint Strike Fighter program, and update of the DoD Human Engineering MIL-STD-1472 and MIL-HDBK-46855. Lt Col. Cunningham (Ret.) has over 29 years of experience in DoD management engineering, aircraft maintenance, reliability and maintainability, acquisition, and logistics. He was director of the Air Force's model HSI organization and managed studies and analyses offices for three AF Major Commands (MAJCOMs) where he performed computer simulations to determine manpower-system design tradeoffs. Mr. Cunningham holds a BS in Industrial Technology and an MS in Systems Management.

Dr Winston Bennett, Jr., as a Personnel Research Psychologist with the AFRL Warfighter Training Research Division, Mesa AZ, is team leader for combat mission analysis, rehearsal, and training technology insertion, and performance assessment R&D. He holds a Ph.D. in Industrial Psychology from Texas A&M University. Recent honors have included the 1998 American Society for Training and Development (ASTD) Research Award and the 1999 AFRL Harry G. Armstrong Scientific Achievement Award. Dr Bennett has published over 50 research articles, book chapters, and technical reports, and presented numerous professional papers related to performance evaluation, personnel assessment, training requirements identification, quantifying the impact of training systems on individual, team, and organizational effectiveness, and statistical methods for cumulating research results.

Using Mission Essential MOEs/MOPs for Evaluating Effectiveness of Distributed Mission Training

Frank C. Gentner, T. Cliff Tiller, and Paul H. Cunningham
University of Dayton Research Institute Human Factors Group
300 College Park, Dayton, OH 45469-0157, U.S.A.

Winston Bennett, Jr.
Air Force Research Laboratory, Human Effectiveness Directorate
Warfighter Training Research Division
6030 S. Kent, Mesa, AZ 85212-6004, U.S.A.

BACKGROUND

Traditional Approaches to Training Evaluation

Typically, traditional approaches to training evaluation have relied on student feedback and measures of learning, such as written and performance tests, and on supervisor feedback on graduates. These methods are easy to implement, inexpensive, but are often not tied to the overall training questions and objectives of the operators in the field or the warfighters making funding decisions. Operators are concerned with interventions and techniques that have a demonstrable impact on their on-the-job behavior and performance. Decision-makers are concerned with providing the most effective training at the lowest cost. In both cases, these individuals need answers to questions regarding the role advanced training systems are to play in total training systems and the impact of such systems on mission effectiveness. They need answers to which tasks can be best trained in mission-simulation, part-task trainers, or require operational flight training in the aircraft. They are also concerned with negative transfer of training, and with task decay and setting recurrent training intervals. They want valid answers to practical questions.

To answer these questions, it is necessary to examine performance from multiple perspectives and with measures that are tied to different aspects of the training system. Tannenbaum and Yukl (1992) recommended that future research designed to systematically evaluate the effectiveness of any training or organizational program must ensure that multiple criteria are used in the evaluation. Even though this recommendation was made some seven years ago, Bennett and Arthur (1999) found very few studies that examined or used multiple criteria at all and even fewer that examined the impact at organizational or mission level. Further, in a training performance measurement

and assessment review covering six military occupational specialties' training, Mohs, MacDiarmid, and Andrews (1988) found a "serious lack of an integrated system for measuring and assessing training performance." Specifically, there was an "over-reliance on subjective measures of performance, and a shortage of valid, reliable quantitative performance measures of training strategies and training effectiveness." (p. i) The same can be said of aircrew training as well.

As training funding decreases and mission requirements increase, the military must be able to directly and quickly determine the impact of training on mission performance. This is especially important when considering the costs associated with high-fidelity flight training simulation systems. Traditionally flight training systems have focused on basic flying skills, but not on the more complex warfighting skills, much less operational team skills, as this quote from the Distributed Mission Training (DMT) Operational Requirements Document (ORD) (Department of the Air Force, 1997) states:

Existing simulation is limited primarily to individual/crew trainers, that do not always reflect latest aircraft configuration and are not designed for interoperable combined exercises. Only basic, single-ship, aircraft training (instruments, emergency procedures, and intercepts) can be accomplished. There is no capability to conduct basic engaged maneuvering and no linkage to allow multiple aircraft to train together to develop complex fighting concepts (large and/or composite force employment) or conduct full mission training in a simulated combat environment at the basic employment formation level. Existing aircrew training capabilities do not support or facilitate realization of COMACC's [Commander of Air Combat Command

(ACC)] vision for high-fidelity distributed mission training. Current training media and the environments in which they are employed do not provide the realism, intensity, or integration required to prepare aircrews to operate effectively on the joint/combined arms battlefield. Aircraft and weapon performance specifications, both friend and foe, are not up-to-date. (Section 3.1)

DMT systems are radical departures from past training simulators. They present revolutionary opportunities not dreamed of only a few years back. These advanced DMT systems offer the promise that they will permit replacement of mission training tasks that are too complex or dangerous to train in an operational environment. Chief of Staff of the Air Force, Gen Ryan, articulated his vision for advanced stimulation systems in his (Department of the Air Force, 1998) posture statement as:

We are also pursuing the development of revolutionary new ways to train our operational aircrews. Distributed mission training will use state-of-the-art distributed simulation technology and advanced flight simulators to permit aircrew to remain at their home units while “flying” and training in synthetic battlespace, hooked electronically to other aircrews located at distant airbases. This will improve the quality and availability of training while reducing aircraft operation and maintenance costs, as well as limiting the amount of time our personnel will have to spend away from home.

Today we stand on the brink of a new frontier having both the need and opportunity to train individual and team warfighting skills more effectively and efficiently. However, we have not advanced our evaluation systems accordingly. We need to determine which aspects of advanced training systems are effective and which need improvement. As simulation systems migrate to focusing on teaching *operational tasks*, should not also training evaluation progress to evaluate *operational performance*, be it in simulations or actual operational performance?

New Operational Methodology for Training Evaluation

Our focus has been the development of an innovative approach to operational evaluation of training. This methodology focuses on linking mission training and rehearsal objectives with operational measures of effectiveness and mission success. Systematically

measuring performance in an operational environment is difficult. Data quality and availability are issues, and for a variety of reasons, application of direct mission performance training evaluation methods has been rare.

Challenges to Operational Evaluation

Approach. A number of factors make operational evaluation of training systems more difficult than traditional training evaluation using schoolhouse measures.

Controlled Environment. First, the lack of a controlled environment presents significant challenges. Establishing a baseline and using students of similar experience is much easier in a school environment where pipeline students, for the most part, have similar experiences. Operational factors consistently interfere with experimental and controlled groups for scientific evaluation. These operational realities include weather, aircraft mechanical performance, operational availability of student pilots and observers who are trained to be consistent, and other compelling mission needs. The post-Cold War period has been characterized by an increased number of conflicts causing higher operational tempo (OPTEMPO) that results in greater interference with planned operational evaluations. Meanwhile in the schoolhouse, the same missions can be flown over and over in a controlled environment with the same caliber of students, and while the same instructors, schooled and practiced in evaluating the same scenario, are nearly always available. It is no wonder that most training evaluation has taken the path of least resistance, especially when the variabilities of the operational setting often make it more difficult to evaluate similar scenarios consistently.

Operational Measures. Until recently, operational measures were only available from a variety of scattered and sometimes incomplete documents, such as checkride checklists, course syllabi, and various regulations (often in draft form) not easily available to training evaluators and researchers. Often evaluators and researchers were not fully aware of the variety of documents to request. Further, the data in these documents did not necessarily reflect the detail or human performance information needed for effective measurement; i.e., few team metrics were addressed. Also, the lack of formal training objectives that reflect higher-order skills like teamwork may contribute to missing measures. Measures in these documents were not hierarchically organized so that they could be discussed, evaluated, and improved. Also, lower-level evaluation criteria were not linked with the higher-level metrics with which warfighter decision-makers are concerned. In the past, a training evaluator paid

attention to the overall mission. Typically, the evaluator could build related scenarios, and could recruit subject matter experts (SMEs – usually instructor pilots) to evaluate those scenarios. However, the consistency among and within raters might be in question because the criteria consisted only of a list of the items to be evaluated and no behaviorally anchored or specific mission success criteria would be designated. In a recent review of pertinent civilian and military literature, regulations, syllabi, and checklists, there was considerable variability across these documents. Different evaluators emphasized different criteria. Some of the most consistent criteria seemed to be the Standardization and Evaluation (STAN/EVAL) criteria recently published in depth by some Air Force Major Commands (MAJCOMs). However, detailed tactics information and criteria have been published in some classified documents, which were not reviewed for this study.

Researcher Awareness of Operational Metrics. Researchers have been concerned with controlled metrics that measure flying skills, since this was the focus of their research and system development. While researchers have been consistently concerned with specific evaluation criteria, many were often not aware of how the criteria “added up” to measure overall operational mission success. They were not focused on the higher level MOEs that MAJCOM commanders and their staffs are concerned with. Therefore, the significance of documented research findings seemed to be microscopic and it was difficult for warfighter decision-makers to understand how training research findings about these small measures could affect mission performance. Aircrew training research, by and large, has been unable to provide these decision-makers with the impact data they need to make mission-related training trade-offs. The bottom line is that if researchers want warfighter decision-makers to listen to their findings, they must use the language and level of operational metrics used by these decision-makers.

Research Findings on Operational Metrics. Given the challenges outlined above, we recently began a study to categorize mission tasks and order officially endorsed metrics into a hierarchical structure. The goal of this effort is to provide a unique capability to target advanced mission-oriented training systems on tasks and measures that are quantifiable and meaningful to field commanders. This taxonomy was to collocate in one document or database tasks that can relate lower-level operational metrics with higher-level MOEs. Given the expense associated with advanced training systems, such as the DMT System, it is crucial to have reliable and valid methods to quantify the

benefits of such systems. Ultimately we want to use the data obtained to develop objective measures that can be easily used by researchers and training evaluators to fine-tune mission training and rehearsal systems while relating training level tasks and MOPs to mission-oriented tasks and MOEs of concern to combat commanders.

Initial Research to Develop MOE/MOP Taxonomy. In our initial efforts, we developed and refined hierarchical taxonomies of operational military performance for aircraft maintainers and operators, as well as other operational specialties. The initial work on maintainers is complete and will be available from a forthcoming AFRL technical report entitled, *Evaluating Aircrew and Maintainer Warfighter Performance in Aeronautical Systems Using Mission-oriented Measures of Effectiveness* (Best, Gentner, Cunningham, Tiller, Schopper, Morris, & Bennett, in press). Some preliminary work was also conducted during 1995-1997 on an aircrew taxonomy, but detailed work was primarily deferred until the Fall of 1998. Results of the initial aeronautical (maintenance and operator) taxonomy were shared with developers of DoD and Service task lists that were being developed in parallel. Research and development (R&D) progressed in stages as the Joint task measurement system and research funding became available.

Joint and Air Force Focus on Mission Tasks and Metrics. As a result of the 1992 Joint Training Review, the Chairman of the Joint Chiefs of Staff (CJCS) directed that the Joint Staff develop and institutionalize a requirements-based training system to better focus DoD's training resources. The first step was to define the required capabilities in terms of mission tasks, conditions, and standards using the Universal Joint Task List (CJCS Manual 3500.04A, 1996). Responding to the CJCS and publication of this manual, the AF embarked on a process of developing and documenting universal mission tasks and metrics, and documenting them in the Air Force Task List (AFTL). The original 1996 draft Aeronautical System-Human Performance MOE/MOP Taxonomy (Best, Gentner, Cunningham, Schopper, & Morris (1997) was given to DoD, Navy, and AF universal task list developers and MAJCOM points of contract (POCs) for inclusion in their task list documents as they added metrics. After an initial draft AFTL that had been formulated to coincide with the Universal Joint Task List (UJTL) (CJCS, 1996), the Air Force (AF) decided to reformat under the AF core competencies, and they have recently been published in AF Doctrine Document (AFDD) 1-1, the AFTL (August, 1998). The AFTL now includes mission

tasks, conditions, and metrics, written at a very general level to be applicable to the AF as a whole. The Air Staff office for operational training (AF/XOOT) recently required that all AF MAJCOMs and units down through the wing-level develop their own unique METL, based on the AFTL, by 15 Jan 99. As these METLs and metrics are developed, they are to be furnished to the Inspector General offices for inclusion in their evaluation process. (AF/XO Message 171257Z AUG 98, MAJCOM METL Development.)

Aircrew MOE/MOP Taxonomy Development. To develop a comprehensive aircrew MOE/MOP taxonomy that could be used to support evaluation of DMT, we expanded the coverage of the Aircrew MOE/MOP Taxonomy. In 1998, the preliminary aircrew taxonomy was reorganized to be consistent with the newly published Air Force Task List (AFTL) (1998) (see Best, et al., in press). This taxonomy furnished much more detailed information than the 1996-7 versions; however, it is not considered in final condition since it is based on a limited number of attack or fighter aircraft (A/OA-10, F-16, F-15C, and F-15E). The taxonomy covers all aircrew-related tasks and includes additional ones not identified in this initial AFTL. Aircrew MOE/MOP Hierarchical Taxonomy is capable of assisting training and mission evaluators in identifying the relationship of lower-level tasks and metrics to higher-level mission tasks and their metrics for fighter aircraft. This taxonomy is an important springboard for studying mission metrics for operational training evaluation. Below, we discuss the sources of the MOEs/MOPs and organization of the taxonomy.

Sources of Aircrew MOEs/MOPs. The taxonomy was constructed using agreement among a number of unclassified military documents that include the following:

- **Military Standard (MIL-STD) 1776A**, *Aircrew Station and Passenger Accommodations* (DoD, 1994) contained a listing of aircrew MOEs/MOPs. More recently, the eight draft Joint Service Aircrew-related Specification Guides (JSSGs) were used (JSSG-2005-2010, 1998).
- **Mission Area Plans (MAPs)** include both MOEs/MOPs and deficiencies. Also by implication, one can develop additional MOEs/MOPs from these deficiencies.
- **Acquisition Documents**, such as the Mission Need Statements (MNSs) and Operational Requirements Documents (ORDs), including the Requirements Correlation Matrix (RCM).
- **Test and Evaluation Master Plans (TEMP) and Test Reports** contain system MOEs/MOPs that can be translated into human performance ones.

- **Readiness Reporting Documents** assisted AF Policy Directive (AFPD) 10-2, *Readiness* (1993), such as the *Status of Resources and Training System (SORTS) reporting readiness in these areas:* (a) personnel, (b) equipment on hand, (c) training, and (d) equipment condition.
- **Universal Joint Task List (UJTL) and other Service Task Lists.** Chairman of the Joint Chiefs of Staff Manual (CJCSM) 3500.04A contains the UJTL (CJCS, 1996). The Army (*Universal Army Task List*, 1996), Navy (OPNAVINST 3500.38, 1996), and Air Force universal task lists were also consulted. The AFTL (1998) had the greatest influence in terms of structure and organization.
- **Operational Readiness Inspections (ORIs)** are required by AFI 90-201, *Inspector General Activities* (SAF/IGI, 1996), which describes the subjects that are to be covered during ORIs. Air Combat Command (ACC), Inspector General (ACC/IGIX, 1996), published a supplement to this AFI that details specific objective grading criteria.
- **AF Aircrew Evaluation Criteria Instructions.** Two of a series of AFIs (11-2 a and f) covering *Aircrew Evaluation Criteria* were reviewed to determine the support for existing MOEs/MOPs, and to add additional appropriate ones.
- **Military Exercises.** Both the AF and the Joint Chiefs of Staff conduct exercises to ensure the force is ready to perform a variety of missions and contingency actions. Exercises use real aircraft and simulated combat situations to test readiness and train war-fighting ability. *Blue* and *Red Flag* are examples of highly visible exercises; however, exercises are numerous and most are classified.
- **Wargames and Simulations** involve replicating warfare without actual combat, often involving computer simulations. Distributed simulation technology is now integrated into many exercises, competitions, and training programs.
- **Military Competitions** are scheduled every year to determine top honors in various mission areas. Competitions appear to be one of the more interesting sources of MOEs/MOPs. For example, ACC's *William Tell* combines competitions for pilot and weapons director accuracy, munitions loading speed and safety, and aircraft maintenance proficiency. Other examples of competitions include the ACC-sponsored, combined-force competitions, *Gunsmoke*; combined mid- and long-range bombing competitions, *Long Shot*; and Air Mobility Command's (AMC's) transportation tanker and airlift *Rodeo*.

Aircrew MOE/MOP Taxonomy Organization. During the Fall of 1998, the aircrew

taxonomy was updated and reorganized to agree with the newly published AFTL. In addition to correlating with the new AFTL material and format, detailed aircrew MOEs/MOPs have been added from AFI 11-2, Aircrew Evaluation Criteria, and other AF instructions. The taxonomy places these MOEs/MOPs into a semi-hierarchical organization. Tasks, MOEs and MOPs are grouped under the following ten categories. The first seven directly relate to Air Force Tasks (AFTs) allowing lower-level MOPs to be tied to higher-level mission tasks and MOEs. Item eight relates tactical employment metrics that can augment evaluation of specific tactics and that might cut across multiple AFTs. Item nine covers other flight performance measures, and item ten covers human performance metrics, such as situation awareness and cognitive workload. The Aircrew Taxonomy outline is presented here:

1. Air and Space Superiority
2. Precision Engagement
3. Information Superiority
4. Global Attack
5. Global Mobility
6. Agile Combat Support
7. Command and Control
8. Tactical Employment
 - 8a General Tactical Measures
 - 8b Air-to-Surface Measures
 - 8c Surface Attack Measures
 - 8d Air-to-Ground Measures
 - 8e Low-Altitude Tactical Navigation
 - 8f Low-Altitude Tactical Formation
 - 8g Killer Scout Metrics
 - 8h Air-to-Air & Air-to-Ground Targeting
9. Other Flight Performance Measures
 - 9.1 Detectability
 - 9.2 Survivability
 - 9.3 Vulnerability
10. Other Human Performance Measures
 - 10.1 Situation Awareness
 - 10.2 Cognitive Workload
 - 10.3 Other

APPROACH

To develop Mission Task-based Metrics for evaluation of DMT we followed the approach of decomposing mission tasks from AF Core Competency tasks down to the level which can analyze individual, intra-team, and inter-team tasks and their related metrics. At this more definitive level, cognitive task and team task analysis can assist in formulating training scenarios and the most appropriate metrics for the mission tasks to be trained. At the same time this detailed analysis assists trainers, the Aircrew-System MOE/MOP Taxonomy maintains a

hierarchical link to higher-level MOEs of concern to warfighter decision-makers. Below is a sample decomposition, which can be used to illustrate how the tasks and their MOEs/MOPs can be helpful in providing prescriptive feedback to trainers and exercise participants, as well as warfighter decision-makers.

Sample Mission Task Decomposition

Data on sample mission task MOE/MOP decompositions are presented to illustrate how a taxonomic approach can help diagnose actual aircrew mission performance of both individuals and teams. This section contains a sample task decomposition using representative AF Task (AFT) 1.1.1.2, *Conduct Defensive Counterair*. This decomposition is of a special case of this task entitled, *Basic Tactical Intercept of Non-maneuvering Target with AWACS Coordination*. It was selected to illustrate a scenario that involves individual and intra-team aircrew tasks, and inter-team coordination with an Airborne Warning And Control System (AWACS) weapons director. Starting from the top and proceeding down the hierarchy, the overall AFTL area for this task decomposition is *Provide Air and Space Superiority* per AF Doctrine Document (AFDD) 1-1, AFTL (1998). The basic precept of *Air and Space Superiority* is one of the core AF competencies to control this domain in order to enhance or secure freedom of action for friendly forces in all geographical environments. The task decomposition (presented in Table 1) focuses downward through the sub-areas of AFT 1.1, *Provide Counterair Capabilities*, and AFT 1.1.1, *Perform Counterair Functions*, and ending in the sub-area AFT 1.1.1.2, *Conduct Defensive Counterair (DCA)*. In addition to the AFTL, Mission Essential Task Lists (METLs) are being developed by each organizational level within the Air Force from the MAJCOM down to wing level. These new attempts to be more specific about each Air Force unit's mission tasks could also help define the tasks more closely; however, the value of Air Force unit METLs in being more descriptive than the AFTL has yet to be established since the program is in its infancy. There is much yet to be researched about this decomposition process since Wing-level METLs were first required in the Air Force on 15 Jan 99 (see AF/XO 1998). If these METLs add to clarity of task, we'll use them as additional criteria; however, if they are repeats of higher-level AFTs, we will skip the METLs and go directly to the Extended Task Decomposition (EDT) using SMEs.

As there are many possible tasks associated with DCA, we have chosen one narrow aspect of DCA, the *tactical intercept*, to define and decompose for this illustration. For this decomposition, we need an operational

scenario. Our example assumes a scenario in which two F-16 aircraft configured with typical air-to-air ordnance loaded on a Combat Air Patrol (CAP) mission in contact with a controlling AWACS aircraft. In the F-16, as in most fighters, the intercept problem involves using the radar to detect a specific target, then using the appropriate intercept geometry to arrive at a position to identify the target and fire weapons.

With structured questions we can elicit from SMEs much additional information regarding task performance requirements for specific mission scenarios. The advantage of this decomposition

taxonomy system being paired with cognitive task analysis is that information on specific task MOPs can be immediately linked to higher-level MOEs that Numbered Air Force (NAF) and MAJCOM commanders are concerned with. Further, the structured information can provide a broader range of MOPs with which to evaluate the advanced DMT and compare its training outcomes with training conducted in the aircraft or other modes. With structured interviews, SMEs can help identify cognitive processes, teamwork and coordination requirements, and can furnish data for predictive and diagnostic training modeling.

TABLE 1. Task Decomposition Hierarchy.

AFT 1	Provide Air And Space Superiority
AFT 1.1	Provide Counterair Capabilities
AFT 1.1.1	Perform Counterair Functions
AFT 1.1.1.2	Conduct Defensive Counterair
[Insert additional Mission Essential Tasks (METs), if appropriate when they become available, to include the following:]	
AF MET 1.1.1.2.X	Conduct Defensive Counterair
ACC MET 1.1.1.2.X	Conduct Defensive Counterair
NAF MET 1.1.1.2.X	Conduct Defensive Counterair (may include additional sub-description)
WING MET 1.1.1.2.X	Conduct Defensive Counterair (+ additional specifics)
SQUADRON MET 1.1.1.2.X	Conduct Defensive Counterair (+ additional specifics)
[Continue decomposition to lower level sub-tasks and MOPs]	
Extended Decomposition Task (EDT) 1.1.1.2.X.X	Conduct Defensive Counterair via a tactical intercept
EDT 1.1.1.2.X.X.X	Conduct Defensive Counterair via a tactical intercept using 4 F-16s
EDT 1.1.1.2.X.X.X.X	Conduct Defensive Counterair via a tactical intercept using 4 F-16s and AWACS assistance with a Visual Identification (VID) Rules Of Engagement (ROE) restriction
Evaluation Criteria	Specific flight tactics, coordination and execution criteria (e.g. offset angles, airspeed, commit criteria, etc.)
Other Performance Analyses (e.g., Cognitive Task Analysis, Teamwork Analysis, Behavioral Anchored Rating Scales, etc.)	Structured interviews of SMEs can help identify cognitive processes, teamwork and coordination requirements, and can furnish data for predictive and diagnostic training modeling.

Technical Challenges & Proposed Solutions

While this hierarchical taxonomic approach shows much promise, many technical obstacles need to be overcome before it can be completed and routinely used in an automated form.

Efficiency & Feasibility. Developing cognitive task analyses of the entire range of aircrew tasks for each possible mission scenario is daunting and seems an almost impossible project. Therefore, one of the first challenges is to map out a strategy that will enable collection of essential information without consuming all available resources before the task is

completed. We, therefore, propose that we use the Aircrew-System MOE/MOP Taxonomy and a series of questions to SMEs to guide selective analysis, provide links to more important MOEs, and capitalize on similar cognitive task analysis outcomes/data for similar tasks. The result of using the taxonomy and strategic cognitive task analysis and team analysis questions is a more efficient task analysis process that makes this project feasible. First by tracking links and interrelationships found in related tasks and by storing the task analysis information in a retrievable database, the work on similar tasks can be used to streamline data collection on succeeding tasks. By having a ready set of MOEs and MOPs to draw from, SMEs can select from a list of accepted MOEs and MOPs, rather than having to invent them.

Linking DMT performance with Field Performance. This effort will provide empirical methods and technologies to link performance in advanced DMT with field performance MOEs/MOPs, and will permit researchers, who may not be fully aware of field performance metrics, to more readily tie their technological interventions to actual mission objectives and requirements. Also this effort will provide a comprehensive, interactive knowledge representation and assessment system that can be used for scenario generation structure, data capture, and data reuse. It provides a link for connecting newly developed tactics with existing ones that will help organizations like the Expeditionary Air Force (EAF) to develop plans and training to counter enhanced enemy force capabilities.

Database for Storing Reusable Aircrew Task Analysis Data. One significant problem faced by training evaluators is that with the myriad of possible scenarios, SMEs must go through a laborious process of developing metrics for each scenario. To expedite this process we propose to build a database that will provide the empirical linkage necessary to build scenarios, to provide measurement metrics for real-time use, and to evaluate the performance of teams within the scenarios. The goal of the database is to permit easy reuse and prompting of SMEs in the development of metrics for new scenarios that are similar to ones already stored in the database. The key will be developing an innovative way of indexing and retrieving similar scenarios for reuse. We believe that the Aircrew Taxonomy can assist with identification of similar scenarios and metrics.

RESEARCH AGENDA

To perfect a more efficient, effective, and operationally relevant method of training evaluation for DMT, the AFRL Warfighter Training Research Division's

research agenda is structured to use innovative approaches. These approaches include mission and task analysis, hierarchy development, cognitive task and team analysis, and automated job-aiding of SME input to developing training scenarios and metrics. Two proof-of-concept innovative research efforts have begun. At the core is an hierarchical representation of mission tasks decomposed to the level necessary to support scenario development. The decomposition will be stored in a database designed to facilitate reuse of mission task and scenario metrics. This reusable database supports two approaches to developing both individual predictive performance measures and models and to develop and improve team metrics for intra-team and inter-team participants in DMT. (See Figure 1 for an overview of the interactions being studied in the two projects.)

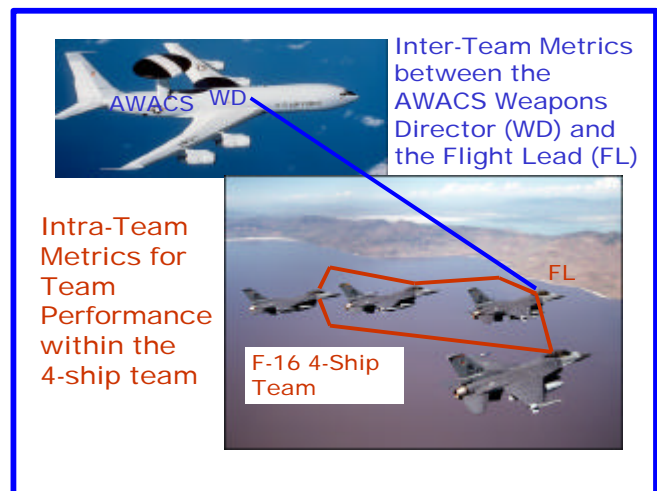


FIGURE 1. Intra- and Inter-Team Interactions Being Studied.

Effort I, Performance Measuring & Modeling Tools

Objective. Our goal in the first project is to develop a proof-of-concept set of tools to specifically tie and model operational measures of performance and effectiveness to training outcomes from the Air Force Research Laboratory's DMT testbed at Mesa, Arizona. It will also provide a means whereby the relevance of instructional strategies or training events to development and sustainment of warfighting knowledge and skill can be demonstrated. The payoff will be to forecast which and how often training events and simulation scenarios are needed to best improve the desired operational outcome.

Approach. This effort will conduct field studies to examine the impact of instructional strategies and

interventions on advanced simulation efficacy and on the development and sustainment of warfighting knowledge and skill with particular focus on the F-16 4-ship pilots and their within-team relationships (see Figure 2, SBIR I). The proposed effort includes examining the utility of an instructor workstation to modify mission simulations and to incorporate instructional strategies into actual simulation events for training purposes. These strategies will have been, a priori, linked to core skill or competencies required for successful mission performance (Mission Essential Competencies). For rapid prototyping, a performance modeling technology similar to the Shipboard Performance Assessment Diagnosis and Evaluation (SPADE), developed for the US Navy (Carolan, Evans, Roth & Scott-Nash, 1997) will be modified and integrated. This effort will include an examination of which metrics can best be obtained from observers and which can best be obtained electronically from the simulator systems. This effort will permit a comprehensive Small Business Innovation Research (SBIR) Phase I demonstration of the proposed suite of tools.

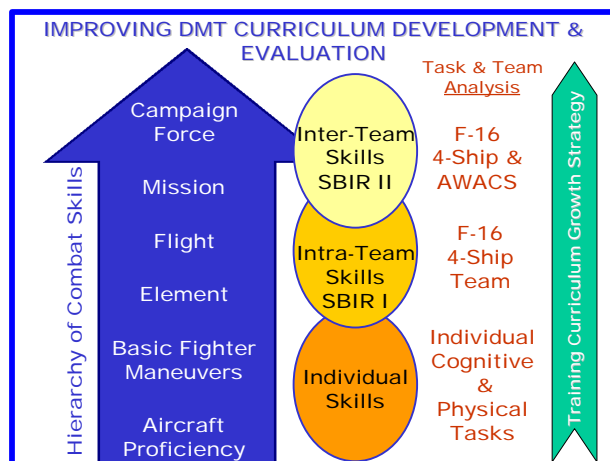


FIGURE 2. Relationship of SBIR Studies to Hierarchy of Combat Skills and Type of Skills to be Analyzed.

Effort II, Team Metric Development and Intervention Tools

Objective. Supported by the interactive mission task decomposition taxonomy database mentioned above, this parallel effort will develop innovative, quantitative, and computer-assisted job performance evaluation methods for individual, workgroup, and team performance and readiness. The focus will be on inter-team metrics, but will not exclude individual and intra-team aircrew metrics. These methods will assess

the impact of manpower, personnel, training, human factors engineering, and knowledge engineering-based interventions on mission performance.

Approach. This Phase I SBIR will develop an innovative set of technological solutions to address key issues in individual and team performance measurement with particular attention to the F-16 4-ship to AWACS inter-team relationship (see Figure 2, SBIR II). This effort will provide empirical methods and technologies to link performance in advanced DMT with field performance and will permit researchers to tie technological interventions to actual mission objectives and requirements. In addition to the team performance measurement technology to be developed, this effort will provide a comprehensive, interactive mission development database that will permit commanders, aircrew, and researchers to develop completely new missions and scenarios that address such things as enhanced enemy force capabilities or new tactics that are developed for the EAF. The database will provide (a) the empirical linkage from similar scenarios necessary to job-aid building new scenarios; (b) assist with building measurement instruments and metrics for real-time use; and (c) assist with evaluation of the performance of teams, as well as individuals, within the scenarios.

Research Plans, Challenges, and Possible Payoffs for MOE/MOP Database

To support these two SBIR projects, the University of Dayton Research Institute Human Factors Group is developing and populating a database to store and relate AFTL mission tasks and metrics with other relevant aircrew performance metrics from the Aircrew Taxonomy (see Figure 3). These tasks and metrics will then be related to mission scenarios, their task elements, and specific MOPs. The research will develop a process for eliciting the MOPs from SMEs and for developing anchored rating scales. These anchored scales can also be stored in the database and catalogued for easy reuse on similar mission scenarios. As you may have detected, one of the challenges and innovations must be a user-oriented logic and cataloging system that enables SMEs to easily select data from existing scenarios that most closely duplicate newly developed scenarios. The payoff of this database is that once these scenarios are selected and duplicated, the SME will only need to change the parts of the scenario task elements and metrics that need revision. Listings of accepted MOPs will be available to choose from. This process could save countless hours of SME time and result in more standard and robust selection of metrics. The long-term goals of the two SBIR projects

are to develop metrics and methods of prescribing interventions to improve performance through using DMT effectively at the individual, intra-team, and inter-team levels. Thus, training interventions must also be linked to the AFTL, expanded tasks, task elements, scenario, and metrics. As our knowledge of this area increases, the database can be expanded and be of even more assistance in developing new mission areas. Two specific challenges are developing the intra-team and inter-team metrics.

Intra-team Metrics. Both individual and within-team (intra-team) tasks steps and their performance metrics (MOPs) must be developed to cover each F-16 4-ship position. The goal is to develop anchored scale metrics that encourage uniformity in rating policy. Predictive models must be developed which aid in diagnosing training deficiencies. To diagnose, one must know appropriate MOPs for both individual and the team. The metrics must be within the context of a scenario to be meaningful; therefore, the scenario must be documented in the database as well. A significant challenge to developing the intra-team metrics will be the process of sorting known mission tasks to subtasks into individual versus team metrics. This is particularly difficult because the metrics are usually written at the collective level in the AFTL and Aircrew Taxonomy. Identification of the individual and team skills will involve detailed SME–Cognitive Task Analyst interaction. Likewise, data to run the predictive models must be extracted from SMEs. Teamwork Knowledge, Skills, and Attitudes (KSAs) are also critical pieces of information (Cannon-Bowers & Salas, 1997), and our plan is to store these in the database, as well.

Inter-team Metrics. So far, we have been unable to find existing metrics for the interaction of the AWACS weapons directors with the F-16 4-ship lead. Thus to develop these MOPs will require similar detailed examination of their interaction with SMEs and cognitive task analyst. The challenge in this process will be to examine AWACS–F-16 Lead interactions to identify team KSAs or competencies. Scenarios must be selected which emphasize sustained interaction between these inter-team (between team) members. By examining the most difficult and critical subtasks, the analyst can zero in on the most relevant metrics for inter-team interaction.

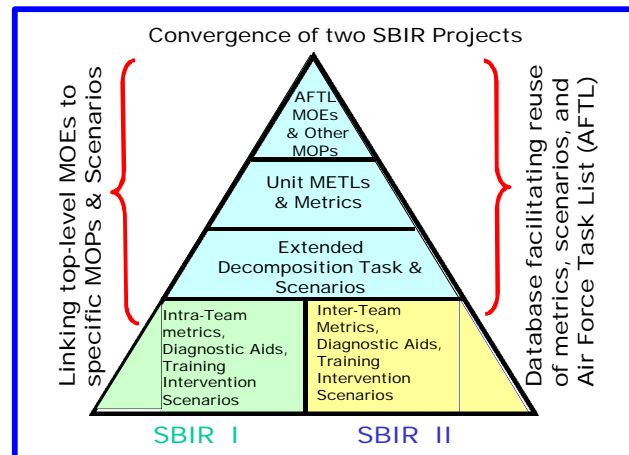


FIGURE 3. Hierarchy of Tasks and Data to be Stored in the Supporting Database.

CONCLUSION

As we develop and organize the individual, intra-team, and inter-team metrics and relate them to the AFTL and other accepted metrics, as well as mission scenarios, this project will provide data needed to begin diagnosis of individual and team performance for prescriptive training interventions. These interventions will then be more operationally relevant by using these proven metrics. Also, when new mission tasks and scenarios are developed because of new technology or new employment strategies, it will be easier to construct valid scenarios with proven metrics. Finally, using robust, validated metrics that are linked to higher-level MOEs will enable decision-makers to more effectively set priorities on the operationally most important training interventions to sponsor.

REFERENCES

- AFDD 1-1, *Air Force Task List*. (1998). Washington, D.C.: Department of the Air Force (AF Doctrine Center).
- AFI 10-201, *Status of Resources and Training System*, (1995). Washington, DC: USAF/XOOOR.
- AFI 11-2F-16, *F-16 Aircrew Evaluation Criteria, Vol 2*. (1998). Washington, D.C.: Department of the Air Force (Air Combat Command).
- AFI 90-201, *Inspector General Activities* (1996).. Washington, DC: Secretary of the Air Force (SAF/IGI).
- AFI 90-201/ACC SUP 1, *Inspector General Activities*. (1996). Langley AFB, VA: Air Combat Command (ACC/IGIX).

- AFPD 10-2, *Readiness*. (1993). Washington, DC: USAF/XOOO.
- AF/XO (1998). AF/XO Message 171257Z AUG 98, *MAJCOM METL Development*. Washington, D.C.: Department of the Air Force.
- Bennett, W., Jr. (1998). Assessing the impact of distributed technology on warfighter performance and mission effectiveness. In Minutes DMT working group Technology Needs Conference. Mesa, AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division.
- Bennett, W., Jr. & Arthur, W., Jr. (in press). *Factors that influence the effectiveness of training in organizations: A review and meta-analysis*(Interim Technical Report). Mesa, AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division.
- Best, P.S., Gentner, F.C., Cunningham, P.H., Schopper, A.W., & Morris, J.D. (1997). *Evaluating war-fighting capabilities of aircraft systems, aircrew, and maintenance using measures of effectiveness and performance*. Wright-Patterson AFB, OH: CSERIAC Program Office.
- Best, P.S., Gentner, F.C., Cunningham, P.H., Tiller, T.C., Schopper, A.W., Morris, J.D. & Bennett, W. (in press). *Evaluating Aircrew and Maintainer Warfighter Performance in Aeronautical Systems Using Mission-oriented Measures of Effectiveness*. Mesa, AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division.
- Cannon-Bowers, J.A., & Salas, E. (1997). Teamwork competencies: The intersection of team member knowledge, skill, and attitudes. In H. F. O'Neil (Ed.) *Workforce Readiness: Competencies and Assessment*. Mahwah, NJ: Erlbaum.
- Carolan, T. F., Evans, D. C., Roth, J. T., & Scott-Nash, S. (1997). Automating Performance Assessment and Diagnosis in Scenario-Based Training. *Proceedings of the Human Factors and Ergonomics Society (HFES)*, 41st Annual Meeting, p. 1123. Santa Monica, CA: HFES.
- CJCS Manual 3500.04A, *Universal Joint Task List*. (1996). Washington, DC: Chairman of the Joint Chiefs of Staff (CJCS).
- Department of the Air Force (1997). Distributed Mission Training Operational Requirements Document (CAF [USAF] 009-93-I-A). Washington, D.C.: Author.
- Department of the Air Force. (1998). *Air Force Posture Statement 1998*. Washington, DC: Author. Available: <http://www.af.mil>.
- Joint Service Specification Guides (JSSG 2005-2010). (1998). Wright-Patterson AFB, OH: Aeronautical System Center Engineering Directorate (ASC/ENSID), Aviation Engineering Board of Joint Aviation Commanders' Group.
- MIL-STD-1776A (1994). Military Standard. *Aircrew Station and Passenger Accommodations*. Washington, DC: DoD.
- Mohs, B., MacDiarmid, W. D., & Andrews, D. H. (1988). *A Review and Annotated Bibliography of Training Performance Measurement and Assessment Literature* (ADA207258). Alexandria, VA: Army Research Institute.
- OPNAVINST 3500.38, *Universal Naval Task List (UNTL) - Section 5: Measures for joint and naval tasks* (1996). Washington, DC: US Navy (OPNAV).
- Tannenbaum S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology*, 43, 399-441.
- Universal Army Task List (UATL)*. (1996). Washington, D.C.: Department of the Army.